



WHAT IS CLAIMED IS:

1. A method of determining a set of large sequences from an electronic data base comprising a set  $D = \{d_1, \dots, d_n\}$  of transactions  $d_i$  ( $1 \leq i \leq n$ ) in a computer system with an implemented query module, each of the large sequences on the set  $D$  of transactions  $d_i$  having a support value greater than or equal to a given support value  $S$ , each of the transactions  $d_i$  of the set  $D$  being a sequence of items of a record  $E = \{e_1, \dots, e_m\}$  of items  $e_j$  ( $1 \leq j \leq m$ ) and the method comprising the following steps:
  - 10 a) determining a set  $L_1$  of large sequences from the set  $D$  of transactions, the large sequences of set  $L_1$  each comprising exactly one item of the record  $E$ , and an assigned support value  $S_{L_1}$  on the sequence  $D$  of transactions each being greater than or equal to the given support value  $S$ ;
  - 15 b) determining a set  $L_2$  of large sequences from the set  $D$  of transactions, the large sequences of set  $L_2$  each comprising exactly two items of the record  $E$  in a respective order  $R_{L_2}$ , and an assigned support value  $S_{L_2}$  on the set  $D$  of transactions each being greater than or equal to the given support value  $S$ , and nothing but sequences comprising one of the large sequences of set  $L_1$ , as a partial sequence, being  
20 taken into account in determining set  $L_2$ ;
  - c) determining a set  $L_k$  ( $k > 2$ ) of large sequences from the set  $D$  of transactions, the large sequences of set  $L_k$  each comprising exactly  $k$  items of record  $E$  in a respective order  $R_{L_k}$ ,  
25 and an assigned support value  $S_{L_k}$  on the sequence  $D$  of transactions each being greater than or equal to the given support value  $S$ , and nothing but sequences comprising two of the large sequences of set  $L_{k-1}$ , as partly overlapping partial sequences, with the respective order  $R_{L_{k-1}}$ , being taken  
30 into account in determining set  $L_k$ ; and
  - d) repeating step c) for  $k = k+1$  and terminating the repetition of step c) when a given termination condition is fulfilled.
2. The method as claimed in claim 1, wherein the set  $D$  of transactions  
35 is searched for candidate sequences which comprise two of

the large sequences of set L1 or set Lk-1, respectively, as partly overlapping partial sequences, in determining set L2 in step b) and set Lk in step c), respectively, and wherein an assigned support value counter is registered for each candidate sequence found for the first time, and the assigned support value counter is incremented when the respective candidate sequence is determined again in searching the set D of transactions.

3. The method as claimed in claim 1, wherein:

- the set D of transactions is searched for candidate sequences comprising one of the large sequences of set L1, as a partial sequence, when determining set L2 in step b), and an assigned support value counter is registered for each candidate sequence found for the first time, and the assigned support value counter is incremented when the respective candidate sequence is determined again in searching the set D of transactions; and
- a candidate set of sequences is formed, when determining set Lk ( $k > 2$ ) in step c), which set comprises all combinations in pairs of the sequences of set Lk-1, and of the set D of transactions those transactions are defined which comprise at least one sequence of the candidate set, and the support value is determined for those sequences of the candidate set for which a transaction was found that comprised this sequence.

4. The method as claimed in claim 2 or 3, wherein a display value is generated and updated for each assigned support value counter to indicate the transaction for which the assigned support value counter was last incremented.

5. The method as claimed in claim 3 or 4, wherein only those sequences of the candidate set of sequences corresponding to all combinations in pairs of the sequences of set L1 are taken into account of which the first item lies within a given range of values.

6. A computer program product for determining a set of large sequences from an electronic data base comprising a set  $D = \{d_1, \dots, d_n\}$  of transactions  $d_i$  ( $1 \leq i \leq n$ ) in a computer system with an implemented query module, each of the large sequences on the set  $D$  of transactions  $d_i$  having a support value greater than or equal to a given support value  $S$ , each of the transactions  $d_i$  of the set  $D$  being a sequence of items of a record  $E = \{e_1, \dots, e_m\}$  of items  $e_j$  ( $1 \leq j \leq m$ ) and the product comprising the following means:

10 a) means recorded on an electronic storage medium for determining a set  $L_1$  of large sequences from the set  $D$  of transactions, the large sequences of set  $L_1$  each comprising exactly one item of the record  $E$ , and an assigned support value  $S_{L_1}$  on the sequence  $D$  of transactions each being greater than or equal to the given support value  $S$ ;

15 b) means recorded on the electronic storage medium for determining a set  $L_2$  of large sequences from the set  $D$  of transactions, the large sequences of set  $L_2$  each comprising exactly two items of the record  $E$  in a respective order  $R_{L_2}$ , and an assigned support value  $S_{L_2}$  on the set  $D$  of transactions each being greater than or equal to the given support value  $S$ , and nothing but sequences comprising one of the large sequences of set  $L_1$ , as a partial sequence, being taken into account in determining set  $L_2$ ;

20 c) means recorded on the storage medium for determining a set  $L_k$  ( $k > 2$ ) of large sequences from the set  $D$  of transactions, the large sequences of set  $L_k$  each comprising exactly  $k$  items of record  $E$  in a respective order  $R_{L_k}$ , and an assigned support value  $S_{L_k}$  on the sequence  $D$  of transactions each being greater than or equal to the given support value  $S$ , and nothing but sequences comprising two of the large sequences of set  $L_{k-1}$ , as partly overlapping partial sequences, with the respective order  $R_{L_{k-1}}$ , being taken into account in determining set  $L_k$ ; and

25 d) means recorded on the electronic storage medium for repeating step c) for  $k = k+1$  and terminating the repetition of step c) when a given termination condition is fulfilled.

7. An integrated sequential analysis system, comprising:

- an electronic data base comprising a set  $D = \{d_1, \dots, d_n\}$  of transactions  $d_i$  ( $1 \leq i \leq n$ ), each of the large sequences on the set  $D$  of transactions  $d_i$  having a support value greater than or equal to a given support value  $S$ , each of the transactions  $d_i$  of the set  $D$  being a sequence of items of a record  $E = \{e_1, \dots, e_m\}$  of items  $e_j$  ( $1 \leq j \leq m$ );
- a query module comprising a query means coupled to the data base and a processing means for detecting query parameters and generating queries to the query means;
- means for determining a set  $L_1$  of large sequences from the set  $D$  of transactions, the large sequences of set  $L_1$  each comprising exactly one item of the record  $E$ , and an assigned support value  $S_{L_1}$  on the sequence  $D$  of transactions each being greater than or equal to the given support value  $S$ ;
- means for determining a set  $L_2$  of large sequences from the set  $D$  of transactions, the large sequences of set  $L_2$  each comprising exactly two items of the record  $E$  in a respective order  $R_{L_2}$ , and an assigned support value  $S_{L_2}$  on the set  $D$  of transactions each being greater than or equal to the given support value  $S$ , and nothing but sequences comprising one of the large sequences of set  $L_1$ , as a partial sequence, being taken into account in determining set  $L_2$ ;
- means for determining a set  $L_k$  ( $k > 2$ ) of large sequences from the set  $D$  of transactions, the large sequences of set  $L_k$  each comprising exactly  $k$  items of record  $E$  in a respective order  $R_{L_k}$ , and an assigned support value  $S_{L_k}$  on the sequence  $D$  of transactions each being greater than or equal to the given support value  $S$ , and nothing but sequences comprising two of the large sequences of set  $L_{k-1}$ , as partly overlapping partial sequences, with the respective order  $R_{L_{k-1}}$ , being taken into account in determining set  $L_k$ ; and
- means for repeating step c) for  $k = k+1$  and terminating the repetition of step c) when a given termination condition is fulfilled.